

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт фундаментальной биологии и биотехнологии  
Кафедра биофизики

УТВЕРЖДАЮ  
Заведующий кафедрой  
\_\_\_\_\_ В.А. Кратасюк

« \_\_\_\_\_ » \_\_\_\_\_ 2020 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Сравнительный анализ характеристик ландшафтов функции  
приспособленности реальных систем и нейросетевых моделей

03.03.02 Физика  
03.04.02.01 Биофизика

Руководитель \_\_\_\_\_ д.ф.-м.н. Барцев С.И.

Выпускник \_\_\_\_\_ Батурина П.М.

Рецензент \_\_\_\_\_ д. б. н. Суховольский В. Г.

Красноярск 2020

## РЕФЕРАТ

Магистерская диссертация по теме «Сравнительный анализ характеристик ландшафтов функции приспособленности реальных систем и нейросетевых моделей » содержит 51 страницу текстового документа, 57 использованных источников, 16 рисунков, 2 таблицы, 16 формул, 2 приложения.

ВОСПРОИЗВОДИМОСТЬ ЭВОЛЮЦИОННЫХ ТРАЕКТОРИЙ,  
ПРЕДСКАЗУЕМОСТЬ ЭВОЛЮЦИИ, СООТНОШЕНИЕ  
ДЕТЕРМИНИРОВАННОСТИ И СТОХАСТИЧНОСТИ ЭВОЛЮЦИИ,  
ПРИНЦИП ОПТИМАЛЬНОСТИ, ЛАНДШАФТ ПРИСПОСОБЛЕННОСТИ,  
ЛАНДШАФТНАЯ НЕРОВНОСТЬ

Тема диссертации связана с основными проблемами эволюционной биологии – детерминированности, воспроизводимости и предсказуемости эволюции. Используется эвристический модельный подход, в качестве объекта которого выбраны формальные нейронные сети, процесс обучения которого феноменологически эквивалентен биологической эволюции. Основная идея работы состоит в сопоставлении свойств ландшафтов приспособленности, (формируемых эволюционными траекториями) реальных систем и нейросетевых моделей. Модельный объект представляет собой полносвязную сеть, обучаемый с помощью алгоритма случайного поиска с возможностью выбора случаев сильного и слабого отбора.

Полученные результаты, демонстрирующие схожие свойства ландшафтов систем обоих типов, открывают перспективы для дальнейших исследований.

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	4
1 Обзор литературы .....	6
1.1 Проблемы детерминированности, воспроизводимости и предсказуемости эволюции	6
1.2 Особенности биологических систем .....	7
1.3 Определение эволюции .....	8
1.4 Эвристическое моделирование .....	9
1.5 Понятие ландшафта приспособленности .....	11
1.6 Особенности исследования ландшафтов приспособленности реальных систем .....	12
1.7 Особенности модельных ландшафта приспособленности .....	15
2 Материалы и методы .....	17
2.1 Сеть формальных нейронов .....	17
2.2 Обучение нейронной сети .....	20
2.3 Параметры обучения нейронной сети .....	21
2.4 Характер дивергентного движения .....	21
2.5 Большие и малые этапы обучения .....	23
2.6 Алгоритм обучения, аналогичный сильному отбору .....	24
2.7 Алгоритм обучения, аналогичный слабому отбору .....	24
3 Результаты и их обсуждение .....	26
3.1 Характерное свойство ландшафта – дивергенция траекторий .....	26
3.2 Характерное свойство ландшафта – кластеры функциональных вариантов .....	28
3.3 Закономерности в формировании кластеров конечных точек .....	30
3.4 Формирование пучков при сильном и слабом отборе. Скорость эволюции. ....	31
3.5 Количественные характеристики ландшафтной неровности .....	34
ВЫВОДЫ .....	38
СПИСОК СОКРАЩЕНИЙ .....	39
Список использованных источников .....	40
Приложение А .....	45
Приложение Б .....	46

## ВВЕДЕНИЕ

Вопрос о том, каково соотношение случайности и детерминированности свойственно процессу биологической эволюции, является актуальным в настоящее время [1–3] и широко исследуется на практическом и теоретическом уровнях.

В первом случае проводят серии экспериментов по молекулярной эволюции для выявления связи между генотипом организма и его приспособленностью [4–6]. Однако возможности эмпирического изучения эволюции ограничены невероятной сложностью изучаемых биологических систем, которым к тому же свойственна в высокой степени уникальность, что ограничивает также возможности применения методов статистического анализа [7–9].

Существующие теоретические модели либо являются высокоспецифичными и сильно параметризованными, либо не учитывают молекулярные и механистические особенности систем [10–14]. В данной работе используется эвристический подход, суть которого в выявлении общих (или типичных) закономерностей широкого класса систем. В качестве модельных объектов эволюции выбраны искусственные нейронные сети, процесс обучения которых сопоставим с процессом реальной эволюции биологических систем.

Работа связана с ключевыми понятиями эволюционных траекторий и ландшафта приспособленности, который они формируют [10]. Особенности формы ландшафта влияют на потенциальные возможности эволюции, а значит на возможности оценки воспроизводимости траекторий и предсказания эволюционных исходов. Поэтому изучение полученных в ходе нейросетевого эвристического моделирования ландшафтов приспособленности, их формы и особенностей является перспективным для понимания эволюции на реальных ландшафтах биосистем.

Целью данной работы является оценка характера формирования эволюционных траекторий и ландшафта приспособленности на примере

нейросетевой эвристической модели и сравнение со свойствами экспериментальных ландшафтов реальных систем.

Для выполнения поставленной цели были поставлены следующие задачи:

- 1) Написать программы для исследования свойств нейросетей в среде SciLab для:
  - Генерации пучков траекторий при сильном отборе из исходной материнской сети и их визуализации.
  - Генерации пучков траекторий при слабом отборе из исходной материнской сети и их визуализации.
- 2) Сравнить полученные в модельных экспериментах данные с данными о реальных системах
- 3) Ознакомиться с публикациями, в которых рассматриваются ландшафты приспособленности реальных и модельных систем, и детально изучить методы количественного описания особенностей ландшафта и используемые меры неровности. Сравнить особенности построения ландшафтов обоих типов и выявить набор характеристик, который можно в дальнейшем использовать для их адекватного сравнения.

## ВЫВОДЫ

- Были написаны программы для исследования свойств нейросетей в среде SciLab для:
  - Генерации пучков траекторий при сильном отборе из исходной материнской сети и их визуализации.
  - Генерации пучков траекторий при слабом отборе из исходной материнской сети и их визуализации.
- Было проведено сравнение полученных в модельных экспериментах данных с данными о реальных системах, в результате которого было показано, что такие особенности нейросетевых ландшафтов, как множественность финальных точек обучения и образование пучков траекторий, к ним приводящих, являются иллюстрацией свойств реальных систем.
- Для дальнейшего, более детального сравнения ландшафтов обоих типов из литературных источников был выявлен набор количественных характеристик неровности.

## **СПИСОК СОКРАЩЕНИЙ**

ЭМО – эвристический модельный объект

ИНС – искусственная нейронная сеть

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Lobkovsky A.E., Wolf Y.I., Koonin E. V. Predictability of Evolutionary Trajectories in Fitness Landscapes // PLoS Comput. Biol. / ed. Shakhnovich E.I. 2011. Vol. 7, № 12. P. e1002302.
2. de Visser J.A.G.M., Krug J. Empirical fitness landscapes and the predictability of evolution // Nat. Rev. Genet. 2014. Vol. 15, № 7. P. 480–490.
3. Bank C. et al. On the (un)predictability of a large intragenic fitness landscape // Proc. Natl. Acad. Sci. 2016. Vol. 113, № 49. P. 14085–14090.
4. Podgornaia A.I., Laub M.T. Pervasive degeneracy and epistasis in a protein-protein interface // Science (80-. ). 2015. Vol. 347, № 6222. P. 673–677.
5. Hayashi Y. et al. Experimental Rugged Fitness Landscape in Protein Sequence Space // PLoS One / ed. Rutherford S. 2006. Vol. 1, № 1. P. e96.
6. Starr T.N., Picton L.K., Thornton J.W. Alternative evolutionary histories in the sequence space of an ancient protein // Nature. 2017. Vol. 549, № 7672. P. 409–413.
7. Hou J. et al. From The Cover: Global mapping of the protein structure space and application in structure-based inference of protein function // Proc. Natl. Acad. Sci. 2005. Vol. 102, № 10. P. 3651–3656.
8. Siltberg-Liberles J., Grahnen J.A., Liberles D.A. The Evolution of Protein Structures and Structural Ensembles Under Functional Constraint // Genes (Basel). 2011. Vol. 2, № 4. P. 748–762.
9. Osadchy M., Kolodny R. Maps of protein structure space reveal a fundamental relationship between protein structure and function // Proc. Natl. Acad. Sci. 2011. Vol. 108, № 30. P. 12301–12306.
10. Gavrillets S. Fitness Landscapes and the Origin of Species (MPB-41). New Jersey: Princeton University Press, 2004.
11. Fragata I. et al. Evolution in the light of fitness landscape theory // Trends Ecol. Evol. 2019. Vol. 34, № 1. P. 69–82.
12. Kauffman S.A., Weinberger E.D. The NK model of rugged fitness



landscapes and its application to maturation of the immune response // J. Theor. Biol. 1989. Vol. 141, № 2. P. 211–245.

13. Poelwijk F.J. et al. Empirical fitness landscapes reveal accessible evolutionary paths // Nature. 2007. Vol. 445, № 7126. P. 383–386.

14. Malan K.M., Engelbrecht A.P. A survey of techniques for characterising fitness landscapes and some possible ways forward // Inf. Sci. (Ny). 2013. Vol. 241. P. 148–163.

15. Gould S.J. Wonderful life: the Burgess Shale and the nature of history. 1st ed. New York: W. W. Norton & Company, 1989. 352 p.

16. Kimura M. The nearly neutral theory of molecular evolution. 1st ed. New York: Cambridge University Press, 1983. 381 p.

17. Eigen M. Selforganization of matter and the evolution of biological macromolecules // Naturwissenschaften. 1971. Vol. 58, № 10. P. 465–523.

18. Мейен С., Соколов Б., Шрейдер Ю. Классическая и неклассическая биология. Феномен Любищева // Вестник АН СССР. 1977. Vol. 10. P. 112–124.

19. Mossio M., Longo G., Stewart J. A computable expression of closure to efficient causation // J. Theor. Biol. 2009. Vol. 257, № 3. P. 489–498.

20. Финкельштейн А., Птицын О. Физика белка. Курс лекций. 1st ed. м: Книжный дом “Университет,” 2002. 376 p.

21. Colombo M. et al. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes // Mol. Ecol. 2013. Vol. 22, № 3. P. 670–684.

22. Rashevsky N. Mathematical Biophysics. The University of Chicago Press, 1948. 669 p.

23. Rashevsky N. Topology and life: In search of general mathematical principles in biology and sociology // Bull. Math. Biophys. 1954. Vol. 16, № 4. P. 317–348.

24. Yu X., Gen M. Introduction to Evolutionary Algorithms. London: Springer London, 2010.

25. Stewart J., Mossio M. Is “Life” Computable? On the Simulation of Closure under Efficient Causation // 4th International Conference on Enactive

Interfaces. Grenoble, 2007. P. 271–276.

26. Фон Нейман Д. Теория самовоспроизводящихся автоматов / ed. Бёрксом закончено и отредактировано А. Москва: Мир, 1971. 382 p.

27. Барцев С.И., Барцева О.Д. Эвристические нейросетевые модели в биофизике: приложение к проблеме структурно-функционального соответствия. Красноярск: Сибирский федеральный университет, 2010. 115 p.

28. Gavrillets S. High-Dimensional Fitness Landscapes and Speciation // Evolution—the Extended Synthesis. The MIT Press, 2010. P. 45–80.

29. Weinreich D.M., Watson R.A., Chao L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories // Evolution (N. Y). 2005. Vol. 59, № 6. P. 1165–1174.

30. Phillips P.C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems // Nat. Rev. Genet. 2008. Vol. 9, № 11. P. 855–867.

31. Poelwijk F.J. et al. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes // J. Theor. Biol. 2011. Vol. 272, № 1. P. 141–144.

32. Lagator M. et al. On the mechanistic nature of epistasis in a canonical cis-regulatory element // Elife. 2017. Vol. 6.

33. Shafee T. Evolvability of a viral protease: experimental evolution of catalysis, robustness and specificity. University of cambridge, 2014.

34. Recent Advances in the Theory and Application of Fitness Landscapes / ed. Richter H., Engelbrecht A. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. Vol. 6.

35. de Visser J.A.G.M., Cooper T.F., Elena S.F. The causes of epistasis // Proc. R. Soc. B Biol. Sci. 2011. Vol. 278, № 1725. P. 3617–3624.

36. Szendro I.G. et al. Quantitative analyses of empirical fitness landscapes // J. Stat. Mech. Theory Exp. 2012. Vol. 2013, № 01. P. P01005.

37. Schenk M.F. et al. Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene // Mol. Biol. Evol. 2013. Vol. 30, № 8. P. 1779–1787.

38. Draghi J.A., Plotkin J.B. Selection biases the prevalence and type of

epistasis along adaptive trajectories // *Evolution* (N. Y). 2013. Vol. 67, № 11. P. 3120–3131.

39. Pumir A., Shraiman B. Epistasis in a Model of Molecular Signal Transduction // *PLoS Comput. Biol.* / ed. Covert M.W. 2011. Vol. 7, № 5. P. e1001134.

40. Chou H.-H. et al. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation // *Science* (80-. ). 2011. Vol. 332, № 6034. P. 1190–1192.

41. Khan A.I. et al. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population // *Science* (80-. ). 2011. Vol. 332, № 6034. P. 1193–1196.

42. DePristo M.A., Weinreich D.M., Hartl D.L. Missense meanderings in sequence space: a biophysical view of protein evolution // *Nat. Rev. Genet.* 2005. Vol. 6, № 9. P. 678–687.

43. Watson R.A., Weinreich D.M., Wakeley J. Genome structure and the benefit of sex // *Evolution* (N. Y). 2011. Vol. 65, № 2. P. 523–536.

44. Vassilev V.K., Fogarty T.C., Miller J.F. Information Characteristics and the Structure of Landscapes // *Evol. Comput.* 2000. Vol. 8, № 1. P. 31–60.

45. Vassilev V.K., Fogarty T.C., Miller J.F. Smoothness, Ruggedness and Neutrality of Fitness Landscapes: from Theory to Application. 2003. P. 3–44.

46. Partridge D., Yates W. Replicability of neural computing experiments // *Complex Syst.* 1996. Vol. 10, № 4. P. 257–282.

47. Горбань А. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей // *Сибирский журнал вычислительной математики.* 1998. Vol. 1, № 1. P. 11–24.

48. Haykin S. *Neural Networks and Learning Machines*. 3rd ed. New Jersey: Pearson, 2009. 936 p.

49. Осовский С. *Нейронные сети для обработки информации*. 2nd ed. Горячая линия-Телеком, 2017. 448 p.

50. Барцев С., Гительзон И., Дегерменджи А. Эвристические

нейросетевые модели и ключевые проблемы биофизики // V Съезд биофизиков России. 2015. P. 11.

51. Bartsev S., Baturina P. Экстремальная и дивергентная компоненты эволюции формируют ландшафт приспособленности эвристических нейросетевых объектов // Нейроинформатика, её приложения и анализ данных. Институт вычислительного моделирования СО РАН, 2019. P. 15–20.

52. Weinreich D.M. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins // Science (80-. ). 2006. Vol. 312, № 5770. P. 111–114.

53. Franke J. et al. Evolutionary Accessibility of Mutational Pathways // PLoS Comput. Biol. / ed. Wilke C.O. 2011. Vol. 7, № 8. P. e1002134.

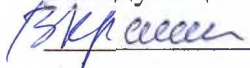
54. Carneiro M., Hartl D.L. Adaptive landscapes and protein evolution // Proc. Natl. Acad. Sci. 2010. Vol. 107, № suppl\_1. P. 1747–1751.

55. Kogenaru M., de Vos M.G.J., Tans S.J. Revealing evolutionary pathways by fitness landscape reconstruction // Crit. Rev. Biochem. Mol. Biol. 2009. Vol. 44, № 4. P. 169–174.

56. Bartsev S., Baturina P. A heuristic neural network model in the research of properties of evolutionary trajectories // IOP Conf. Ser. Mater. Sci. Eng. 2019. Vol. 537. P. 042001.

57. Aita T., Iwakura M., Husimi Y. A cross-section of the fitness landscape of dihydrofolate reductase // Protein Eng. Des. Sel. 2001. Vol. 14, № 9. P. 633–638.

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт фундаментальной биологии и биотехнологии  
Кафедра биофизики

УТВЕРЖДАЮ  
Заведующий кафедрой  
 В.А. Кратасюк  
« 24 » июня 2020 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Сравнительный анализ характеристик ландшафтов функции  
приспособленности реальных систем и нейросетевых моделей

03.03.02 Физика  
03.04.02.01 Биофизика

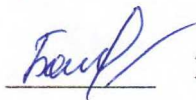
Руководитель



24.06.2020

д.ф.-м.н. Барцев С.И.

Выпускник



24.06.2020

Батурина П.М.

Рецензент



24.06.2020

д. б. н. Суховольский В. Г.

Красноярск 2020